however, masks a somewhat larger (and statistically significant) decline in inferential comprehension (see Table III-6). As defined in the NAEP, inferential comprehension can be considered the highest-level skill tapped by the test. It entails comprehending ideas that are not explicitly stated by drawing inferences from material that is explicit. 38/ In contrast, literal-comprehension scores changed by only a trivial amount, and reference skills--also a more basic area--actually improved, albeit by a very small and statistically insignficant amount.

Deterioration of higher-level skills is also apparent from declines on tests that are designed specifically to tap them. 39/ The SAT is the most salient example. As noted earlier, it is designed (and is generally considered) to be a test that relies heavily on skills such as reasoning, problem-solving abilities, and verbal relationships (such as are assessed by analogies). The Illinois Decade Study (which used a test that was also developed by the Educational Testing Service) provides another example. While the Decade Study included many items that required that students know specific pieces of information (such as rules of English usage, social-studies facts, and mathematical terminology), it also relied heavily on inference. 40/ The declines on the test were relatively large (see Table III-4 and Appendix C).

The relationship between age and the size of the decline--discussed in Chapter IV--might also be indirect evidence of a lesser deterioration of more basic skills. As noted earlier, declines in the first three grades tended to be slight and short-lived and might best be seen as brief interruptions of an otherwise steady upward trend in those grades. Since the curriculum in

---

38. NAEP, *Three National Assessments of Reading*, pp. 4, 25.

39. The tests noted here are all multiple-choice format. As noted in Chapter 2, some people have argued that multiple-choice tests are demonstrably limited in their ability to tap many higher-order skills (for example, Norman Frederiksen, "The Real Test Bias: Influences of Testing on Teaching and Learning," *American Psychologist*, vol. 39 (March 1984), pp. 193-202). Even if the tests noted here leave many relevant higher-order skills unassessed or inadequately measured, however, few people would argue with the notion that they do rely substantially on some higher-order abilities and that those abilities play a greater role in determining scores in these tests than in some others (such as the NAEP literal comprehension reading subtest or the NAEP mathematics test as a whole).

40. Illinois State Board of Education, *Student Achievement in Illinois, 1970 and 1981*, Appendix A.

TABLE III-6. NAEP READING CHANGES 1970-1979, AGE 17, BY AREA (Average percent of items correctly answered)

Area	1970	1979	Change
Total	68.9	68.2	-0.7
Literal Comprehension	72.2	72.0	-0.2
Inferential Comprehension	64.2	62.1	-2.1[a/]
Reference Skills	69.4	70.2	0.8

SOURCE: NAEP, *Three National Assessments of Reading: Changes in Performance, 1970-80,* Table 6.

a. Statistically significant, *p* less than .05.

the early grades includes a large amount of basic skills--decoding and literal comprehension in reading, memorization of basic arithmetic facts, learning of the simplest arithmetic algorithms, spelling, and so on--the almost uninterrupted progress in those grades might reflect relatively favorable trends in the mastery of those particular basic skills.

The Subsequent Upturn. The characteristics of the subsequent upturn are as yet less clear, in part because scores on tests administered in high schools began improving only recently. That the upturn is occurring in most tests and at all grade levels--including the SAT in the last few years--suggests that improvements are probably occurring at many skill levels, but there is as yet little direct indication of the relative size of the upturn in different types and levels of skills. Moreover, the pattern may be complex; for example, the upturn may have different components in different grade levels or among different groups of students.

Disquieting but incomplete suggestions of relatively smaller increases in higher-level skills are found in the most recent (1981-82) NAEP mathematics assessment. Because the NAEP tests a nationally representative

sample of students and because it permits comparisons of changes in various skill areas, it is a particularly important indicator of the mix of skills comprised by recent trends. The NAEP found a sizable increase in the performance of 13-year-olds between 1977-1978 and 1981-1982 (but no appreciable change in the performance of 17-year-olds and only slight and statistically insignificant gains among 9-year-olds). 41/ The nature of the improvement among 13-year-olds, however, was disturbing:

> ...They improved most on the knowledge, skills, and understanding exercises, and least on the applications exercises. Further study shows that their improvements in understanding came on exercises judged relatively easy by a panel of mathematics educators; performance levels on exercises calling for *deeper* understanding showed little or no improvement. 42/

On the other hand, recent gains among the highest-achieving students on difficult tests--discussed in the following chapter--suggest improvement in their higher-order skills. It is possible that some groups of the highest-achieving students are gaining substantially in higher-order skills, while many other students are showing less progress in this regard, but available data remain too limited to answer this question.

41. The lack of change among 17-year-olds, but not the absence of substantial improvement among 9-year-olds, is predictable on the basis of the cohort model discussed earlier.

42. NAEP, *Third National Mathematics Assessment*, p. xv.

CHAPTER IV

GROUP DIFFERENCES IN ACHIEVEMENT TRENDS

While the achievement decline and the subsequent upturn occurred among most groups of students identifiable in the existing data, both trends varied among different groups. Similarly, achievement trends have varied among different types of communities and schools.

The most important differences in trends are:

- Greater declines on tests administered to older students;
- Relative gains by black and Hispanic students, compared with nonminority students; and
- Relative gains in high-minority schools and schools in disadvantaged urban communities compared with the nation as a whole.

In addition, there is some indication that students in the bottom fourth of the achievement distribution gained ground relative to those in the top fourth during part of the 1970s. The evidence on this point is inconsistent, however, and it is not clear that this narrowing of the gap occurred on a variety of tests or spanned more than a short period of time. Female students also showed slightly sharper declines on language-related tests (such as reading and vocabulary), but not on tests in other subject areas. Private school students showed declines comparable to those among public school students in reading and vocabulary, although evidence from a single test suggests that the decline in mathematics achievement was considerably smaller among private-school students.

DIFFERENCES IN TRENDS AMONG TYPES OF STUDENTS

Variation in achievement trends were associated with age, sex, achievement subgroup (that is, low versus high achievers), and race and ethnicity.

Age

Both the decline in achievement and the subsequent upturn varied markedly with the age of the students tested, but the effects of age appear to have been different during the two periods.

The Decline. The total size of the decline was strongly related to age. In general, tests administered to older students showed markedly larger total declines than did tests administered in the early grades. 1/

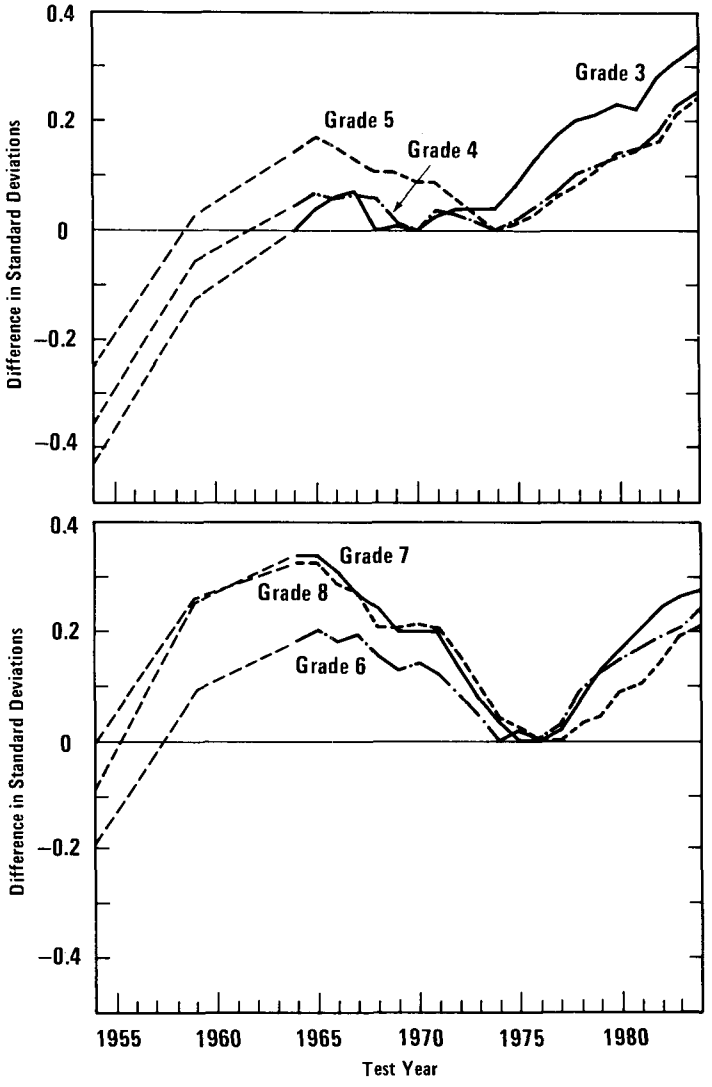
The Iowa state data provide the best assessment of this question and show a striking link between age and the size of the achievement decline (see Figures IV-1 and IV-2). 2/ At one extreme, the decline in third-grade scores was small and short-lived; it can be characterized as a slight dip accompanying an eight-year hiatus in an otherwise unbroken, 30-year increase in achievement. In standardized form, the total decline was only about 0.07 standard deviation (depending on subject), and average scores are now over a third of a standard deviation above the low point of the decline--and more than three-fourths of a standard deviation above their level of

---

1. Although this conclusion is widely accepted, it is important to note that it is actually based on fairly limited data. To offer a good test of the relationship between age and the size of the decline, a data series should meet a set of criteria that few do. The data series should include comparable tests administered to a range of ages, since a comparison of different tests can confound differences between the tests themselves with the effects of age. Scores should be presented in some form--such as standard deviations or percentiles--that permits comparisons among grades. The data should also extend back to the onset of the decline. Data that extend over a relatively short period of time might tap a relatively steep portion of the decline in one grade and a relatively gradual portion in another, thus biasing the comparison among age groups. In addition, random year-to-year fluctuations in scores--reflecting either sampling fluctuations or uncontrolled differences in tests--are more likely to bias conclusions based on a relatively few years. Finally, the data should be annual, to confirm that they subsume the entire decline and none of the upturn. Data that are collected intermittently--such as the NAEP and norming data from commercial elementary and secondary tests--can mix in varying periods of increasing scores for different age groups. Intermittent data also might capture a relatively steep portion of the decline in one grade but a comparatively gradual portion in another.

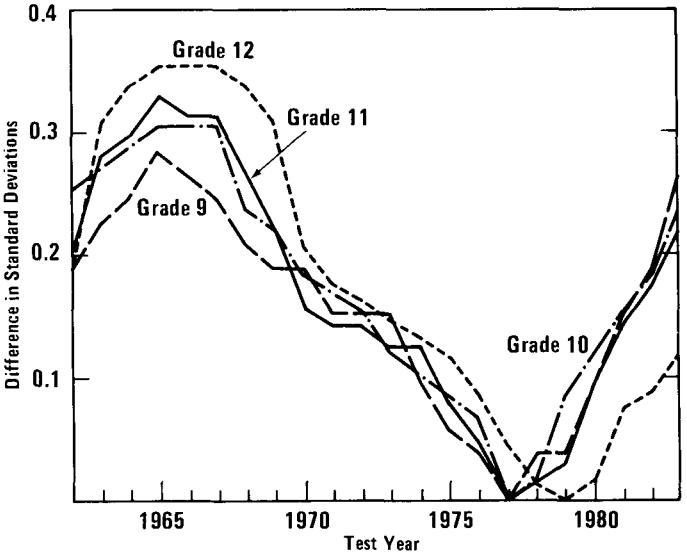
2. The best assessment of the effect of age is obtained within each test series--that is, comparing ITBS scores in grades 3 through 8 with each other, and similarly comparing ITED scores in grades 9 through 12. Even in this case, comparisons across the two tests --for example, comparing grade 8 ITBS scores with the grade 9 ITED--confounds differences between the two tests with the effects of age. (See the discussion in Chapter III of differences in trends among subject areas for a concrete example of differences of this sort between the ITBS and ITED.)

Figure IV-1.
Iowa Composite, ITBS, Grades 3-8, Differences from Post-1964 Low Point



SOURCES: CBO calculations based on "Iowa Basic Skills Testing Program, Achievement Trends in Iowa: 1955-1985" (Iowa Testing Programs, unpublished and undated material); and A. N. Hieronymus, E. F. Lindquist, and H. D. Hoover, *Iowa Tests of Basic Skills: Manual For School Administrators* (Chicago: Riverside, 1982).

Figure IV-2.
Iowa Composite, ITED, Grades 9-12, Differences from Lowest Year



SOURCES: CBO calculations based on "Mean ITED Test Scores by Grade and Subtest for the State of Iowa: 1962 to Present" (Iowa Testing Programs, unpublished and undated tabulations); Robert Forsyth, Iowa Testing Programs, personal communication, August 1984.

three decades ago. Thus, the median third-grader in Iowa today scores better than roughly 68 percent of his or her counterparts of three decades past. Similarly, no sizable decline occurred in grade three in statewide assessments in New York and California. 3/

The decline in eighth-grade Iowa scores, in contrast, was large enough to depress composite achievement scores to their level of three decades ago and long enough that recovery has as yet been incomplete. When put in standard form, these differences appear even more striking. Eighth-grade Iowa scores declined about a third of a standard deviation and have since

3. New York State Education Department, unpublished tabulations; Frank Armbruster, Paul J. Bracken, and Joyce Lind, *The U. S. Primary and Secondary Educational Process* (Croton-on-Hudson, New York: The Hudson Institute, 1975), Appendix A; Dale Carlson, California State Department of Education, personal communication, March 1984.

recovered only about two-thirds of what they lost. (Nonetheless, eighth-grade scores are still about 0.2 standard deviation higher than they were 30 years ago, placing the median student this year at the 58th percentile relative to achievement levels in 1954.)

The National Assessment of Educational Progress (NAEP) also shows only relatively few and small declines among nine-year-olds, relative to the declines in the older groups. This pattern might in part reflect the timing of the NAEP assessments, however, rather than--or in addition to--truly lesser declines in the youngest age group. 4/

Periodic national norming data from commercial standardized elementary and secondary tests also suggest both a lack of decline in the youngest age groups and progressively larger declines in the remainder of the school-age population. For example, the national ITBS norming data indicate that in reading, the median third-grader's level of achievement increased by 4.3 months from 1955 and 1963, only 0.5 months from 1963 to 1970, and 3.7 months from 1970 to 1977. This change is consistent with the pattern in the annual Iowa data--that is, a pause in achievement growth in the late 1960s and early 1970s. In contrast, among sixth graders, a 2.2-month gain from 1955 to 1963 was followed by declines of 2.6 and 3.0 months in the following seven-year periods. Among eighth graders, the drop was even more substantial after 1970. 5/ The SRA achievement series showed composite gains in all but one grade between 1962 and 1971. Between 1970 and 1977, however, the trends varied greatly with grade level. In reading, for example, the latter period included large gains (two-thirds of a standard deviation or more) in grades one and two; more moderate gains in grades three and four; small declines in grades five through eight; and larger drops in the higher grades. 6/

---

4. Given the cohort pattern shown by the end of the decline, the various NAEP assessment cycles probably began near or even at the end of the decline among nine-year-olds, and thus the data most likely combine a few years of the decline with a longer period of the subsequent increase. Since the NAEP assessments are conducted only at intervals of four or five years, however, the precise end of the decline in that test series cannot be firmly established, and the extent of this confounding therefore cannot be determined.

5.. A. N. Hieronymus, E. F. Linquist, and H. D. Hoover, *Iowa Test of Basic Skills: Manual for School Administrators* (Chicago: Riverside Publishing Company, 1982).

6. Science Research Associates, *SRA Achievement Series, Technical Report #3* (Chicago: SRA, 1981), Table 2; and Science Research Associates, unpublished tabulations. The trends between the 1970 and 1977 school years reported here reflect normings conducted in the springs of 1971 and 1978 and are labeled in terms of those calendar years in the published data.

Although the achievement decline persisted longer in the higher grades, the larger total drop in scores in those grades reflects more than the longer duration. In addition, the decline appears to have been steeper--that is, more rapid--in the higher grades. This rapidity is shown most clearly by the Iowa data (both the ITBS and the ITED; see Figures IV-1 and IV-2). In all but two cases, the decline in any grade was steeper than that in all lower grades. This difference in the rapidity of the decline, however, appears to have been confined primarily to the earlier years of the decline.

The Upturn. As noted in Chapter III, scores on tests administered to younger children have risen substantially more in recent years, compared with the decline in those grades, than have scores on tests administered in the higher grades. This pattern can be seen clearly in the Iowa state trend data (both the ITBS and ITED; see Figures III-2 and III-3):

- Grades 3, 4, 5, and 6 are now at their highest point in the three decades of available data.

- Achievement in grades 7, 8, 9, and 10 has rebounded strongly but is not yet at its earlier high (although grade 9 is nearly at that level).

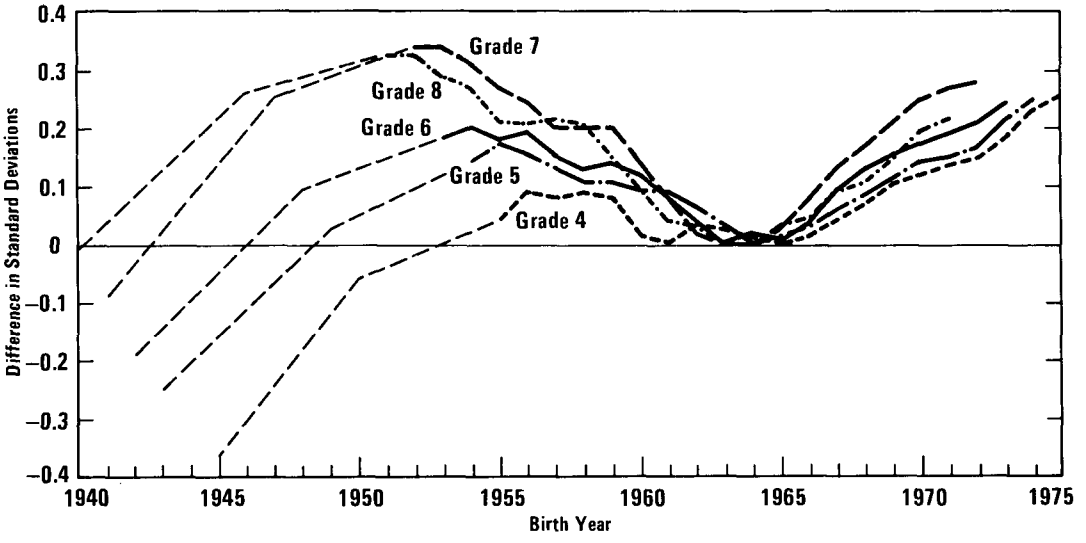
- Grade 12 achievement has begun rising but remains near its low point.

The well-known SAT trend parallels the twelfth grade Iowa trend in this regard: achievement has been climbing for several years but remains only modestly above its low point (see Figure III-4). Similar patterns--although often less clear-cut--appear in a number of other data bases as well, such as the Virginia State assessment data and the NAEP reading assessment. (Some achievement test series, however, are inconsistent with this pattern. For example, in the NAEP mathematics assessment, the recent increase in performance was markedly greater among 13-year-olds than among 9-year-olds.) 7/

The greater total rise in scores to date in the younger grades appears largely to reflect a longer period of rising scores in those grades rather than a greater rate of improvement than in the higher grades. The upturn in scores followed quickly after the end of the decline and shows the same

7. National Assessment of Educational Progress, *The Third National Mathematics Assessment: Results, Trends, and Issues* (Denver: NAEP/Education Commission of the States, 1983), Table 5.1.

Figure IV-3.
ITBS Composite, by Birth Year



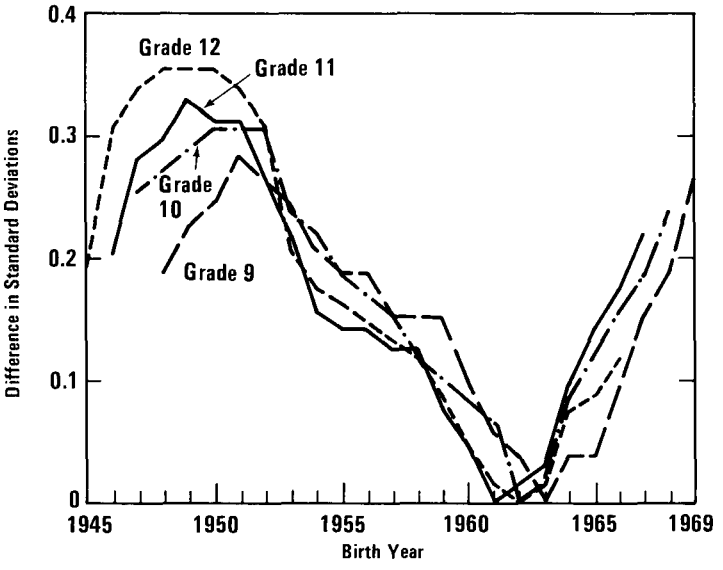
SOURCES: CBO calculations based on "Iowa Basic Skills Testing Program, Achievement Trends in Iowa, 1955-1985" (Iowa Testing Programs, unpublished and undated material); and A. N. Hieronymus, E. F. Lindquist, and H. D. Hoover, *Iowa Tests of Basic Skills: Manual For School Administrators* (Chicago: Riverside, 1982).

cohort pattern (see Appendix B). Among children born after 1963 or so--that is, beginning with the cohorts that entered school in the late 1960s--each cohort has tended to outscore those preceding it. The smaller gains in the higher grades thus appears to reflect, at least in substantial part, the smaller number of higher-scoring cohorts that have reached senior high school.

This trend can be seen in the Iowa data, which suggest--if trends in Iowa are indicative of national trends in this regard--that gains have been comparably fast, or even more rapid, in the higher grades than in the lower ones.[8/] On the ITBS, each birth cohort since the onset of the score increase has tended to produce slightly larger increases in grades six through eight than in grades four and five (see Figure IV-3; vertically adjacent lines that

8. This conclusion reflects changes expressed in standard deviations and only comparisons within a single test. Trends on the ITED are not compared with those on the ITBS.

Figure IV-4.
ITED Composite, by Birth Year



SOURCES: CBO calculations based on "Mean ITED Test Scores by Grade and Subtest for the State of Iowa: 1962 to Present" (Iowa Testing Program, undated and unpublished tabulations); Robert Forsyth, personal communication, August 1984.

are parallel indicate comparable gains by the same cohort in different grades). In this respect, the upturn in the ITBS has been largely symmetrical with the last years of its downturn. On the ITED, the gains produced by any given cohort have remained roughly comparable as that group moved from grade 9 through grade 12 (see Figure IV-4). For several cohorts after the upturn began, these gains were also basically symmetrical with the corresponding last years of the decline, but the most recent cohorts to reach the high-school years--those born in 1966 through 1969--have produced gains that are larger than the corresponding decline produced by the birth cohorts of the mid-1950s.

Sex

While the achievement decline was sizable among students of both sexes, it was somewhat more severe among female students in the case of language-related tests (such as vocabulary, reading, and the SAT-Verbal). On the other hand, once the effects of changes in the composition of the test-

taking group are taken into account, the declines among males appear to have been comparable or even slightly larger than those among females in mathematics and science. 9/

The average SAT scores of women dropped substantially more than those of men. This difference by sex was large on the verbal scale--after 1967, women dropped 50 points, compared with the 36-point drop in the average score of males--but far smaller on the mathematical scale. 10/ The average score of female ACT candidates also dropped more than that of males, and the difference was greater on the English test than in mathematics. 11/

In both cases, however, the apparently greater decline among women might simply be a reflection--at least in part--of the changing mix of male and female students taking the tests. Women have constituted a growing share of all students taking both the SAT and the ACT. Women constituted 42.7 percent of SAT candidates in 1960, 47.5 percent in 1970, and 51.8 percent in 1983. 12/ Similarly, women constituted 45 percent of ACT candidates in 1964 and 54 percent both in 1975 (the year that ACT scores

---

9. On the ACT, the greater decline among women was most pronounced in social studies. In the NAEP, however, the only comparison in social studies that showed relatively greater trends in one gender than the other--citizenship questions at age 13--showed females gaining relative to males. Comparable tabulations from other tests are unavailable. The sharp decline of women on the ACT social studies test therefore might be just a reflection of the compositional changes discussed below. L. A. Munday, *Declining Admission Test Scores*, Research Report #71 (Iowa City: American College Testing Program, February 1976); National Assessment of Educational Progress, *Changes in Political Knowledge and Attitudes*, 1969-76 (Denver: NAEP/Education Commission of the States, March 1978.)

10. College Entrance Examination Board, *College-Bound Seniors, 1984* (New York: The College Board, 1984).

11. These patterns reflect changes in ACT scores from 1965 to 1975, the latter being the year in which composite ACT scores reached their lowest point. The data from 1965 to 1969 are slightly inconsistent with the later data because the former include residual on-campus testing. The former are taken from Munday, *Declining Admission Test Scores*; the latter are from unpublished ACT tabulations.

12. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination* (New York: College Entrance Examination Board, 1977), p. 16; and College Board, *College-Bound Seniors, 1984*.

reached their low point) and in 1983. 13/ This growing share suggests that the pool of women taking the tests might have become relatively less select--a change that would lead to greater score declines among women than among men.

Trends in scores on other tests, however, suggest that part of the greater decline among women is independent of these compositional changes, reflecting some other, as yet unidentified, factors. Data from a few nationally representative tests--which are largely free of these compositional changes--also show greater declines among female students on language-related tests. 14/ On the other hand, in mathematics and science the decline in the scores of male students was typically as large or even larger. For example, a comparison of the high-school classes of 1972 and 1980 found that women showed a greater decline in vocabulary and a slightly larger drop in reading, while men showed a larger decline in mathematics. 15/ Seventeen-year-olds showed a similar pattern in the NAEP over a five- to nine-year span in the 1970s. Women showed a greater decline on both the literal comprehension and inferential comprehension components of the reading assessments, while men evidenced slightly greater declines in mathematics and science. 16/ Although these differences by sex were very small, they might have been larger if the comparisons had spanned the entire period of the achievement decline rather than only a portion of it.

---

13. Munday, *Declining Admissions Test Scores*; American College Testing Program, unpublished tabulations.

14. Although nationally representative data are most often largely free of this particular type of compositional change, they are not always entirely devoid of it. For example, data based on high school samples could show a change of this sort if trends in dropout rates differed markedly by sex.

15. Donald A. Rock, Ruth B. Ekstrom, Margaret E. Goertz, Thomas L. Hilton, and Judith Pollack, *Factors Associated with Decline of Test Scores of High School Seniors, 1972 to 1980* (Washington, D.C.: Center for Statistics, U.S. Department of Education, 1985).

16. National Assessment of Educational Progress, *Three National Assessments of Reading* (Denver: NAEP/ Education Commission of the States, 1981), Tables A-9, A-10, and A-11; *Mathematical Technical Report: Summary Volume* (Denver: NAEP/ Education Commission of the States, 1980), Table 4; *Three National Assessments of Science: Changes in Achievement, 1969-77* (Denver: NAEP/ Education Commission of the States, 1978), Table A-4. In the case of science, the scores of women increased, while those of men dropped.

Achievement Subgroups

A current and widely held view is that the decline in achievement was more severe among relatively high-achieving students than among those at the lower end of the achievement distribution. This belief has led some observers to credit the educational system with improving its services to low-achieving students, or, alternatively, to fault it for allowing its services for more able students to deteriorate. 17/

It is not clear, however, that trends have been consistently more favorable among lower-achieving than among higher-achieving students over the entire period of the achievement decline and subsequent upturn. When a wide range of tests is considered, a more complex--and sometimes inconsistent--pattern emerges. Moreover, there are major gaps in the available data--such as the sparseness of relevant comparisons during the first half of the achievement decline, and a very limited picture of the relative performance of achievement subgroups during the recent years of increasing achievement. In addition, both apparent changes in the gap between achievement subgroups and inconsistencies in the data about these groups must be taken cautiously because both consistencies and variations in the data can be artifacts of technical aspects of the tests.

As discussed in Chapter II, a number of technical aspects of tests influence conclusions about relative trends in high- and low-achieving groups. Differences in the scaling of test scores can markedly affect such judgments. In addition, a single test is unlikely to be a comparably comprehensive measure of mastery at two very different levels of achievement and therefore may understate the relative change of students at one level. The tabulation and reporting of results further complicates comparisons, since information on the additional items correctly or incorrectly answered is rarely reported, particularly for achievement subgroups. This lack of information makes it hard to judge whether changes in the average scores of achievement subgroups are substantively comparable, even when they seem similar numerically. Nonetheless, the broad range of tests suggests the following generalizations. (See Appendix D for additional details.)

---

17. See, for example, statement by Archie E. Lapointe, Executive Director, National Assessment of Educational Progress, before the Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor, January 31, 1984; and William W. Turnbull, *Changes in SAT Scores: What Do They Teach Us?* (report to the College Board-ETS Joint Staff Research and Development Committee, forthcoming).

It is clear that the achievement decline and the subsequent upturn appeared among both low- and high-achieving students. Whether the decline began at the same time in different achievement subgroups, however, and whether the drop was comparable among those subgroups during the early years of the decline (the late 1960s and the first years of the 1970s) remain unknown. Tabulations comparing achievement subgroups during those years are largely restricted to unrepresentative groups of students--for example, comparisons of students taking the SAT, classified in terms of their rank on that test.

During the mid- and late 1970s--that is, during the end of the achievement decline and the beginning of the subsequent upturn--students in the top achievement quartile (the top fourth of all students, when ranked by achievement) lost ground relative to those in the bottom quartile in reading, mathematics, and science in the National Assessment of Educational Progress. That pattern appeared in all three age groups tested (ages 9, 13, and 17), although it took different forms at different ages--probably as a result of the cohort pattern shown by the end of the decline. At age nine, gains predominated over losses, but the lowest quartile showed larger gains than did the highest. At age 17, declines predominated, with the larger losses generally appearing in the highest quartile. At age 13, gains and losses were more evenly mixed, but the lowest quartile still showed relative gains.

While the narrowing of the gap between the top and bottom achievement quartiles on the NAEP is clear-cut, other data cast doubt on the extent to which this was a general trend over the past two decades. Similar trends appear in some data (such as the Illinois Decade Study and some tabulations of the SAT), but not on others (such as other tabulations of the SAT).[18/] Moreover, under most circumstances, a narrowing of the gap between the top and bottom quartiles would cause the standard deviation of test scores--that is, their variability--to decrease. That has not been the general pattern, however, in the few data sources for which historical records of standard deviations are available. Since the early 1970s, the standard deviations (SDs) of the SAT and ACT have been stable or increasing slightly. The SD of the ITBS has been increasing, while that of the SRA achievement series has shown mixed trends (generally inconsistent with

18. The Illinois Decade Study is a comparison of the performance of Illinois high school juniors on a fairly high-level battery of achievement tests in the 1970 and 1981 school years. See Appendix D.

the NAEP pattern in the earlier grades, but consistent in the higher grades). 19/

Several explanations of this inconsistency are plausible. Some of the variation among tests could simply be an artifact of scaling differences. For example, the Illinois Decade Study is consistent with the NAEP in its published form, which presents simple differences in scores, but is inconsistent when presented in terms of proportional changes in scores. Differences in the way students are classified as high- and low-achieving could also account for much of the variability. For example, classifying students in terms of their self-reported class rank yields patterns on the SAT since 1975 that are consistent with the NAEP (even though the standard deviation of the SAT was increasing at that time), while classifying students in terms of their rank on the SAT itself yielded trends that are inconsistent with the NAEP. On the other hand, some of the inconsistency might reflect true variation among tests; perhaps the lowest quartile gained relative to the highest only on certain types of tests.

Test scores of students taking college admissions tests--currently, about half of all high school graduates--declined more than those of high school seniors in general. But this difference primarily reflects the changing composition of the group taking those tests rather than a greater decline in achievement among high-achieving students. The proportion of students taking the SAT, for example, grew substantially during the 1960s and early 1970s, and this growth was accompanied by an increase in the share of SAT candidates from historically lower-achieving groups, such as certain ethnic groups and families of lower socio-economic status.20/ Since the early 1970s, however, such changes in the composition of the test taking group have been relatively minor. 21/

The highest-achieving students--those scoring highest on tests, taking the most advanced courses, and so on--evidenced both the decline and the subsequent upturn in achievement. These students did not show a consistently greater decline than the average student. Indeed, by some measures,

---

19. The College Board, *College-Bound Seniors*, various years; American College Testing Program, unpublished tabulations; H.D. Hoover, personal communication, March 1984; and Science Research Associates, *SRA Achievement Series, Technical Report #3*, Table 2.

20. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination.*

21. Because compositional changes exacerbated the decline in the SAT but not the subsequent upturn, comparing the SAT upturn to the previous decline is misleading. The relative size of the upturn is understated unless adjustments are made to compensate for the compositional changes.

they appear to have gained recently relative to the average, particularly in the area of mathematics. For example, the proportion of SAT candidates scoring over 700 on the mathematics test has risen sharply in the last few years (from 2.7 percent in 1980 to 3.6 percent in 1984) and is now quite close to the level of 1966--the highest level in any year for which tabulations are available. Similarly, American seniors taking calculus and pre-calculus--together about 10 percent to 12 percent of all seniors--showed gains between 1964 and 1981 in international assessments of mathematics achievement. The sketchiness and inconsistency of data on the highest-achieving students, however, cloud these conclusions.

RACE AND ETHNICITY

Recent years have seen a shrinking of the long-standing difference between the scores of black and nonminority students on a variety of achievement tests. The evidence pertaining to other ethnic groups is more limited, but there are suggestions of relative gains by Hispanic students as well.[22] While the change has been small relative to the remaining gap between the minority and nonminority students, it has been consistent from year to

22. The term "ethnicity" as used in the following discussion encompasses some distinctions --such as that between blacks and whites--that are often popularly termed racial. This convention is followed in part for simplicity, but also because some of the most common current categories have at best ambiguous racial bases. For example, many South Asians are often classified as nonwhite (as in some Census tabulations), even though most South Asians are in fact racially Caucasian. Similarly, people of mixed black/white origin are frequently classified as black without regard to whether the greater proportion of their ancestry is in fact white or black. Hispanics are almost all classified as whites in Census tabulations, even though many of them are racially mixed. (In particular, many are partially or primarily native American in origin, and native Americans are racially classified as "Mongoloid"--that is, Asian--people.)

    The ethnic categories used in this paper necessarily reflect the disparate conventions used in the data sources cited and therefore vary among tests. In general, the term "nonminority" excludes, to the extent possible, all minority groups identified in each data source and usually corresponds to the category labeled "white" in the cited sources. The data sources vary considerably, however, in terms of how many--and which--groups are specifically identified. Moreover, some individuals--such as black Hispanics--can be classified in more than one way, and there is typically little information available about how those ambiguous cases are handled.

    The more important known variations in the classifications used in the various sources are noted in Appendix E.

year and could prove substantial over the long run. These patterns are summarized below and are discussed in more detail in Appendix E.

Trend data on the scores of different ethnic groups are very limited, however, and generally extend back only a relatively short length of time. In addition, since many ethnic-group differences in achievement are large, the ambiguity inherent in measuring changes in the gaps between achievement subgroups described above applies to these comparisons as well. In this case, however, the pattern of the trends leaves no doubt that the closing of the gap is at least in part real and not an artifact of the tests.23/ Finally, classification of students' ethnicity is likely to be prone to error, both because of the unreliability of students' self-reports and because of the ambiguity--and lack of consistency over time--of ethnic classifications. While this is unlikely to be a serious source of bias in interpreting trends among black students, it is cause for caution in considering data about Hispanics. 24/

Black Students. In general, it appears that the average scores of black students:

- Declined less than those of nonminority students during the later years of the general decline;
- Stopped declining, or began increasing again, earlier; and
- Rose at a faster rate after the general upturn in achievement began.

---

23. This narrowing of the gap is substantiated by several factors. First, the pattern is consistent among a variety of very different tests. Second, during certain periods, the convergence reflected gains among blacks concurrent with declines among nonminority students. Unlike differences in relative gains (or declines) between groups, a pattern of gains in one group and declines in the other is unlikely to be an artifact of the scaling method used and will generally persist even if the data are rescaled. Third, biases caused by ceiling effects have been largely ruled out. In the case of tests scored as the percent of questions answered correctly, the scores of the higher-achieving group can be held down by a ceiling effect, creating an illusion that lower-achieving groups are gaining in comparison. To lessen the likelihood of such a distortion, data of that sort were transformed (by a logit transformation) to eliminate ceilings, and the narrowing of the gap remained.

24. See, for example, "Problems in Defining Ethnic Origin," Appendix A in Congressional Research Service, *Hispanic Children in Poverty* (Washington, D.C.: CRS, September 13, 1985).